Evidence procedure for efficient quantum state tomography

Jochen Rau*

Institut für Theoretische Physik, Johann Wolfgang Goethe-Universität,
Max-von-Laue-Str. 1, 60438 Frankfurt am Main, Germany
(Dated: June 29, 2010)

I show that in tomographic experiments measurement of a small set of observables suffices to confirm or incrementally amend prior expectations with a high degree of confidence. To this end I adapt the evidence procedure, an estimation technique used in classical image reconstruction, to use in quantum state tomography.

I. INTRODUCTION

The reconstruction of a quantum state from experimental data is generally imprecise for a variety of reasons. Measurement devices work with limited accuracy; the sample size is finite; and the set of observables used might not be informationally complete. Rather than singling out a quantum state uniquely, such imperfect data are generally compatible with an infinite set of different states. Identifying the most plausible member of this set can be a formidable challenge that requires advanced statistical estimation techniques [1]. For this task there exist numerous theoretical schemes, some widely used examples being pattern functions [2, 3], maximum likelihood [4–6] or – in case of informational incompleteness – the maximum entropy method [7–10].

Yet while, e.g., maximum likelihood estimation has proven successful in many practical applications, it reaches a theoretical limit and may even lead to inconsistencies when applied to special situations: For instance, it may return zero eigenvalues for a density matrix when, in fact, this is just an artefact caused by a small sample size [11]. This suggests that prior knowledge beyond the "naked" data may be important, and has motivated Bayesian modifications to the conventional scheme [12]. Another shortcoming of many algorithms is that they do not quantify error bars, a piece of information that would be crucial to assess the reliability of an estimate; here, too, important theoretical developments are currently underway [13, 14].

In addition to their conceptual limitations most reconstruction schemes run into practical difficulties as soon as quantum systems become more complex, as the requirements for experimental and computational resources grow exponentially with the number of constituents of a composite system. For example, Häffner et al. [15] report that, to reconstruct an entangled state of eight calcium ions, they needed to perform 656,100 experiments. This has triggered an intense search for alternative quantum tomography protocols that can do with fewer resources [16].

Recently a number of proposals have been made for efficient protocols whose resource requirements grow only polynomially with system size. All these proposals are limited in scope: Either they do without complete state reconstruction and instead focus on ascertaining some specific property of the state such as the presence or absence of entanglement [17, 18]; or else they presuppose that the unknown state lie in some lower-dimensional reconstruction subspace [19] or belong to some privileged subclass of states such as matrix product states, with a number of parameters which is only polynomial in system size [20–22]. Protocols of the latter type deliver one or both of the following: (i) the information whether or not an unknown state does indeed belong (within some given error bound) to the privileged subclass; and (ii) if so, an estimate for the associated state parameters. While this falls short of tomography for arbitrary states it is nevertheless of great use for many real-world situations as, e.g., matrix product states capture the low energy physics of a wide range of one-dimensional systems [22].

The recent proposals illustrate that efficiency gains can be achieved if one makes clever use of *prior knowledge* about the system — such as in the above example, the well-founded expectation that the system is described by a matrix product state. Prior knowledge thus plays a pivotal, and often underappreciated, role in quantum state tomography. Its use in a tomographic scheme not only ensures internal consistency and in particular avoids the zero-eigenvalue problem encountered in conventional maximum likelihood; but it is also key to realizing sorely needed efficiency gains. In most experimental settings such prior knowledge is in fact available, in the form of well-founded (implicit or explicit) expectations as to the output state. These expectations might be based on a theoretical model of the underlying physics, past results of the same or similar experiments, or some combination thereof; and they may specify an anticipated output state uniquely, or merely its parametric form.

The situation where prior expectations merely favor a certain subclass of states, without any bias as to the parameter values within this subclass, can in general be reduced to the case where one anticipates some unique output state. For as long as there is a polynomial-time algorithm for finding within the subclass the member which best matches some given experimental data – as

* jochen.rau@q-info.org; www.q-info.org

is the case for matrix product states – one can always pre-process the data to obtain the best match within the subclass, and subsequently consider this best match to be the unique anticipated state. In many realistic cases, therefore, one may assume that prior knowledge comes in the form of some unique anticipated state.

While there is thus an expectation as to the output state, it is obviously not guaranteed that precisely this anticipated state will be produced in the experiment — after all, if this were certain a priori, what would be the point of the experiment? The anticipated state comes only with a finite (and often not precisely quantified) degree of confidence: It is well possible that due to unwarranted approximations in the theoretical model or experimental inaccuracies the actual output state will deviate from the anticipated state. This deviation is not expected to be very large, for else there would be a serious problem with the theory or experimental setup; but a quantitative bound is difficult to establish a priori. Strictly speaking, therefore, the prior expectation is that the actual state lies within some unspecified, yet not too large neighborhood of the anticipated state. Often the precise magnitude of this deviation constitutes one of the novel insights provided by the experiment.

In this generic setting quantum state tomography reduces to the following tasks: (i) verify that the unknown state of the system is indeed in the proximity of (but not necessarily identical with) the anticipated state; (ii) whenever it deviates from the anticipated state, prescribe how this initial estimate must be amended; and (iii) quantify the error bars associated with the updated estimate. The recent proposals mentioned above accomplish these tasks in part: They do supply efficient protocols, yet only for the special case where the anticipated state is of matrix product form, and only for the first of the three tasks. Against this backdrop, the purpose of the present paper is to propose a universal protocol with a significantly broader scope: a protocol which takes into account prior knowledge in a generic way, allowing arbitrary anticipated states, and which tackles efficiently all three tasks at once.

In order to achieve this objective I will borrow powerful techniques from the classical field of image reconstruction [23, 24]. There, too, a large amount of noisy (pixel) data needs to be processed with sophisticated algorithms to yield the most plausible tomographic image. Like in quantum state tomography, one key to improving the efficiency and accuracy of these algorithms has been to use whatever prior knowledge is available, however qualitative or vague. Mathematically, this is achieved by modelling prior knowledge with distributions that contain themselves unknown parameters; these so-called *hyperparameters* are then themselves subjected to an estimation procedure. In its most general form such a scheme is known as *hierarchical Bayes*. Practical calculations usually involve well-controlled approximations, the most important being the *evidence procedure* [25–33]. It entails a maximum likelihood approximation applied to the estimation of hyperparameters and is therefore also called *generalized maximum likelihood* or *ML-II*.

Transferring this idea to quantum state tomography, the unknown degree of confidence as to the anticipated state becomes a hyperparameter. This hyperparameter can only be estimated a posteriori based on the overall compatibility of experimental data with the anticipated state. The estimation involves a generalized maximum likelihood approximation which, as it pertains to a one-dimensional distribution only, is usually of good quality — more so than the conventional maximum likelihood approximation which pertains to a higher-dimensional, and hence generally broader, distribution on state space. The thus estimated hyperparameter establishes whether or not the unknown state of the system is indeed in the proximity of the anticipated state; and if so, it determines the optimal weight to be attributed to anticipated state and experimental data, respectively, yielding a posterior estimate that interpolates between the two. The procedure is optimal in the sense that it maximizes the degree of confidence about the resulting posterior estimate; and it is efficient in that it does so with high accuracy even when measurements are not informationally complete.

In the present paper I flesh out these ideas in mathematical detail and demonstrate their use in a concrete example. I shall first adapt the framework of the classical evidence procedure to its novel use in quantum state tomography. I will establish the criteria for its applicability and show that it is in fact particularly suited for tomography of large quantum systems. I will further show that the evidence procedure is internally consistent and in particular avoids the zero-eigenvalue problem encountered in conventional maximum likelihood; that it greatly speeds up tomography (certification or incremental amendment of an arbitrary anticipated state) whenever the experimental data are sufficiently close to prior expectations overall; and that it provides error bars in a straightforward fashion. I will then illustrate its use in a simple four-qubit system.

For simplicity I shall disregard inaccuracies introduced by the experimental apparatus and instead focus on the two other sources of imprecision, finite sample size and lack of informational completeness. Incorporating experimental errors into the formalism is a straightforward exercise that will be dealt with in future work.

In Section II I introduce the evidence procedure, already in a mathematical form adapted to quantum state tomography. In Section III I illustrate its use with a toy example, a four-qubit system with parameters tuned (for simplicity) such that all calculations can be done analytically. In Section IV I conclude with a brief discussion.

II. EVIDENCE PROCEDURE

Suppose one wants to infer the state ρ of a d-dimensional quantum system from a set of r sample means $\{g_a\}$ measured on a sample of size N. Since N

is always finite, and the set of observables $\{G_a\}$ need not be informationally complete, these data do not specify ρ uniquely but only yield some probability distribution $\text{prob}(\rho|N,\{g_a\})$. If this distribution is sharply peaked at some ρ_e then this ρ_e constitutes a plausible estimate for ρ; and the width of the peak is a measure for the quality of the estimate. Due to Bayes' rule the post-measurement distribution is the product of likelihood and prior,

$$\text{prob}(\rho|N,\{g_a\}) \propto \text{prob}(\{g_a\}|N,\rho) \cdot \text{prob}(\rho). \quad (1)$$

The conventional method of maximum likelihood disregards the prior and focuses on the peak of the likelihood function. In contrast, the evidence procedure (as do other Bayesian methods) seeks to encode in the prior whatever ancillary information is available, however qualitative and vague, and use this to improve the accuracy and efficiency of the estimate.

Often prior knowledge comes in the form of a bias towards some anticipated state σ derived from theoretical considerations or past experience. If this is the only information available then one expects the prior to be peaked at, and symmetric around, σ; yet the degree of confidence as to this bias, and hence the width of the distribution, are generally not known. Based on very general considerations of classical statistical inference [34] which readily carry over to the quantum case, the prior must be a monotonically decreasing function f of the distance of ρ from σ, measured in terms of the relative entropy:

$$\text{prob}(\rho) = f[S(\rho\|\sigma)], \quad (2)$$

where $S(\rho\|\sigma) := \text{tr}(\rho\ln\rho - \rho\ln\sigma)$. In the evidence procedure the prior is in fact modelled as an integral

$$\text{prob}(\rho) = \int_0^\infty d\alpha\,\text{prob}(\rho|\alpha) \cdot \text{prob}(\alpha) \quad (3)$$

where

$$\text{prob}(\rho|\alpha) = Z(\alpha)^{-1}\exp[-\alpha S(\rho\|\sigma)] \quad (4)$$

with (in the Gaussian approximation) $Z(\alpha) \approx (2\pi/\alpha)^{n/2}$ and $n = d^2 - 1$. The prior thus features an unknown *hyperparameter* α whose value is to be estimated a posteriori.

Given the experimental data the distribution for α is updated according to Bayes' rule

$$\text{prob}(\alpha|N,\{g_a\}) \propto \text{prob}(\{g_a\}|N,\alpha) \cdot \text{prob}(\alpha). \quad (5)$$

While the prior $\text{prob}(\alpha)$ is generally broad, reflecting the vagueness of the initial bias, the likelihood function may be peaked around some α_0 at which

$$\left.\frac{\partial}{\partial\alpha}\right|_{\alpha_0}\text{prob}(\{g_a\}|N,\alpha) = 0. \quad (6)$$

Whenever this peak is sufficiently sharp,

$$-\alpha_0^2\left.\frac{\partial^2}{\partial\alpha^2}\right|_{\alpha_0}\ln\text{prob}(\{g_a\}|N,\alpha) \gg 1, \quad (7)$$

it is justified to make the approximation

$$\text{prob}(\rho) \to \text{prob}(\rho|\alpha_0). \quad (8)$$

This amounts to a maximum likelihood approximation which, in contrast to the conventional approach, no longer pertains to the estimation of the full quantum state but to that of the hyperparameter α – whence the method's alternative names *generalized maximum likelihood* or *ML-II*.

Whenever the approximation (8) is justified state estimation can be based on the posterior

$$\text{prob}(\rho|N,\{g_a\},\alpha_0) \propto \text{prob}(\{g_a\}|N,\rho) \cdot \text{prob}(\rho|\alpha_0). \quad (9)$$

Thanks to the quantum Stein lemma [35–38] the likelihood function can for large N be written in the form

$$\text{prob}(\{g_a\}|N,\rho) \propto \exp[-NS(\mu_g^\rho\|\rho)] \quad (10)$$

where μ_g^ρ is the unique state that minimizes the relative entropy $S(\mu\|\rho)$ with respect to ρ while yielding as expectation values of $\{G_a\}$ the observed sample means $\{g_a\}$. In the Gaussian approximation it is

$$S(\mu_g^\rho\|\rho) \approx S(\rho\|\mu_g^\rho) \approx S(\mu_{g(\rho)}^\sigma\|\mu_g^\sigma) \quad (11)$$

where $\mu_{g(\rho)}^\sigma$ and μ_g^σ now minimize the relative entropy with respect to the initial bias σ rather than ρ, while yielding for $\{G_a\}$ the same expectation values as ρ, $\{g_a(\rho)\}$, or the observed sample means $\{g_a\}$, respectively. The various μ's all have a generalized canonical form [39]: e.g.,

$$\mu_g^\sigma \propto \exp\left[(\ln\sigma - \langle\ln\sigma\rangle_\sigma) - \sum_{a=1}^r \lambda^a G_a\right] \quad (12)$$

with Lagrange parameters $\{\lambda^a\}$ adjusted such that the state yields as expectation values of $\{G_a\}$ the observed sample means. For a typical bias towards total ignorance, $\sigma = I/d$, this state coincides with the estimate based on maximum entropy [9].

The second factor in the posterior (9) is the prior (4). Like the likelihood function it features in the exponent a relative entropy, which can be decomposed into

$$S(\rho\|\sigma) = S(\rho\|\mu_{g(\rho)}^\sigma) + S(\mu_{g(\rho)}^\sigma\|\sigma). \quad (13)$$

Summing up the various terms in the exponents of likelihood and prior, and exploiting the quasi-linearity

$$(1-t)S(\rho\|\mu) + tS(\rho\|\sigma) = S(\rho\|\rho(\mu,\sigma;t)) + C(\mu,\sigma;t) \quad (14)$$

where $\rho(\mu,\sigma;t) \propto \exp[(1-t)\ln\mu + t\ln\sigma]$ and C is independent of ρ, the posterior finally acquires the form

$$\text{prob}(\rho|N,\{g_a\},\alpha_0) \propto \exp[-\alpha_0 S(\rho\|\mu_{g(\rho)}^\sigma) - (\alpha_0 + N)S(\mu_{g(\rho)}^\sigma\|\rho_e)]. \quad (15)$$

It is peaked around

$$\rho_e \propto \exp\left[\frac{\alpha_0}{\alpha_0 + N}\ln\sigma + \frac{N}{\alpha_0 + N}\ln\mu_g^\sigma\right], \quad (16)$$

with error bars of the order $O(1/\sqrt{\alpha_0 + N})$ as to the degrees of freedom that have been measured, and $O(1/\sqrt{\alpha_0})$ as to those that have not. The estimate ρ_e has the same parametric form as the (generalized) Maxent estimate (12), yet with Lagrange parameters rescaled by a factor $N/(\alpha_0 + N)$. The evidence procedure thus interpolates between the initial bias and the experimental data: For $N \ll \alpha_0$ quantum state estimation is dominated by the prior, and one is advised to stick to the initial bias. For small sample sizes the zero-eigenvalue problem is thus avoided. At the opposite extreme, $N \gg \alpha_0$, the rescaling factor approaches unity; hence as one would expect, quantum state estimation is then dominated by the likelihood function (10) and leads to the (generalized) Maxent estimate (12).

The optimal value α_0 for the hyperparameter is yet to be determined, and the validity of the (generalized) maximum likelihood approximation yet to be established. As for the former, one introduces the marginalization

$$\text{prob}(\{g_a\}|N,\alpha) = \int d\rho \, \text{prob}(\{g_a\}|N,\rho) \cdot \text{prob}(\rho|\alpha) \tag{17}$$

and uses

$$(\partial/\partial\alpha)\text{prob}(\rho|\alpha) = [n/2\alpha - S(\rho\|\sigma)] \cdot \text{prob}(\rho|\alpha) \tag{18}$$

to derive from the extremum condition (6) an implicit equation for α_0:

$$\langle \alpha_0 S(\rho\|\sigma)\rangle_{\alpha_0} = n/2, \tag{19}$$

where $\langle \ldots \rangle_{\alpha_0}$ denotes the expectation value calculated with the posterior (15). In the Gaussian approximation this expectation value can be decomposed into three summands

$$\langle \alpha_0 S(\rho\|\sigma)\rangle_{\alpha_0} = \langle \alpha_0 S(\rho\|\mu^\sigma_{g(\rho)})\rangle_{\alpha_0} + \langle \alpha_0 S(\mu^\sigma_{g(\rho)}\|\rho_e)\rangle_{\alpha_0} + \alpha_0 S(\rho_e\|\sigma) \tag{20}$$

yielding, respectively,

$$\langle \alpha_0 S(\rho\|\mu^\sigma_{g(\rho)})\rangle_{\alpha_0} = (n-r)/2, \tag{21}$$

$$\langle \alpha_0 S(\mu^\sigma_{g(\rho)}\|\rho_e)\rangle_{\alpha_0} = \alpha_0/(\alpha_0 + N) \cdot r/2, \tag{22}$$

$$\alpha_0 S(\rho_e\|\sigma) = \alpha_0 N^2/(\alpha_0 + N)^2 \cdot S(\mu^\sigma_g\|\sigma), \tag{23}$$

provided the r sample means are independent. This leads to the explicit formula

$$\alpha_0 = (1 - N_{\min}/N)^{-1} \cdot N_{\min} \tag{24}$$

with

$$N_{\min} := r/2 S(\mu^\sigma_g\|\sigma). \tag{25}$$

In order for the evidence procedure to be applicable the hyperparameter must be non-negative, and hence the sample size must exceed $N_{\min}$. In addition, the inequality (7) must hold. By a similar reasoning as above the latter translates into an inequality for the variance of the relative entropy,

$$\text{var}(\alpha_0 S(\rho\|\sigma)) \ll n/2 - 1, \tag{26}$$

the variance again being evaluated in the posterior (15). As before, in the Gaussian approximation this variance can be decomposed into three summands

$$\text{var}(\alpha_0 S(\rho\|\sigma)) = \text{var}(\alpha_0 S(\rho\|\mu^\sigma_{g(\rho)})) + \text{var}(\alpha_0 S(\mu^\sigma_{g(\rho)}\|\rho_e)) + 4\alpha_0^2 \langle ([\sigma - \rho_e]; [\rho - \rho_e])^2_{\rho_e}\rangle_{\alpha_0} \tag{27}$$

where in the last summand $(\cdot; \cdot)_{\rho_e}$ denotes the scalar product induced by the entropy [40, 41], evaluated at ρ_e. These three terms yield, respectively,

$$\text{var}(\alpha_0 S(\rho\|\mu^\sigma_{g(\rho)})) = (n-r)/2, \tag{28}$$

$$\text{var}(\alpha_0 S(\mu^\sigma_{g(\rho)}\|\rho_e)) = \alpha_0^2/(\alpha_0 + N)^2 \cdot r/2, \tag{29}$$

$$4\alpha_0^2 \langle ([\sigma - \rho_e]; [\rho - \rho_e])^2_{\rho_e}\rangle_{\alpha_0} = 2\alpha_0^2 N^2/(\alpha_0 + N)^3 \cdot S(\mu^\sigma_g\|\sigma). \tag{30}$$

Inserting for the hyperparameter the formula (24) then gives the criterion

$$r/2 \cdot (1 - N_{\min}/N)^2 \gg 1; \tag{31}$$

i.e., the evidence procedure presupposes that the set of observables is large, $r \gg 1$. Thus the evidence procedure is adapted in particular to the tomography of quantum systems that are large.

Combining the last inequality once more with the formula (24) for the optimal hyperparameter shows that the latter must be inside the range

$$N_{\min} < \alpha_0 \ll \sqrt{r/2} \cdot N_{\min} \tag{32}$$

and hence approximately of the same size as $N_{\min}$. Consequently, fluctuations in hitherto *unmeasured* degrees of freedom are effectively estimated to be of the order $O(1/\sqrt{N_{\min}})$; the hitherto unmeasured degrees of freedom are estimated to an accuracy *as if* they had been measured on a sample of size $N_{\min}$. This fictitious sample size $N_{\min}$ becomes large whenever $S(\mu^\sigma_g\|\sigma)$ is small, i.e., whenever the experimental data are in good agreement with prior expectations overall. In this case the data, albeit informationally incomplete, provide a high degree of confidence not only as to the degrees of freedom actually measured but also as to those that have not. If one seeks confirmation or only incremental amendment of prior expectations, therefore, it is not necessary to measure a set of observables that is informationally complete.

III. EXAMPLE: FOUR QUBITS

In this Section I apply the evidence procedure to incomplete tomography of a four-qubit system. The example is intentionally simplified in order to make all calculations tractable analytically.

An M-qubit system (say, M entangled photons) is described in a Hilbert space of dimension $d = 2^M$. Complete state tomography would require measurement of $n = 2^{2M} - 1$ different observables, i.e., demand resources that grow exponentially with the number of qubits. As M increases, this quickly becomes impossible in real experiments. Here I suppose that in an actual experiment one can only determine sample means $\{\vec{s}_i\}$ of the single-qubit spin vectors

$$\vec{S}_i := 1/2 \cdot (X_i, Y_i, Z_i) \qquad (33)$$

where X_i, Y_i, Z_i are the Pauli matrices pertaining to qubit i, and sample means $\{\mathbf{c}_{ij}\}$ of the qubit-qubit correlation matrices with components

$$C_{ij}^{ab} := S_i^a S_j^b - s_i^a s_j^b , \ a, b = 1 \ldots 3 , \ i \neq j. \qquad (34)$$

Together these are $r = 3M^2$ independent sample means, a number growing only quadratically with the number of qubits. For four qubits it is $r = 48$, less than 20% of an informationally complete set ($n = 255$), but still much greater than one and hence within the range where one can apply the evidence procedure.

Before collecting experimental data and applying the evidence procedure one must reveal one's initial bias σ. In this example I assume that there is no prior information available about a specific order of the qubits, so σ must exhibit the symmetry

$$\langle \vec{S}_i \rangle_\sigma = \vec{s}(\sigma) \ \forall i , \ \langle C_{ij}^{ab} \rangle_\sigma = c^{ab}(\sigma) \ \forall i \neq j. \qquad (35)$$

Moreover, there shall be no reason a priori to prefer any particular spatial direction, an indifference which must be reflected in the initial bias being isotropic:

$$\vec{s}(\sigma) = 0 , \ c^{ab}(\sigma) = c(\sigma) \cdot \delta^{ab}. \qquad (36)$$

Yet I do assume that – based on either theoretical considerations or past experience with similar experimental setups – one has reason to believe that the qubits are correlated, and hence $c(\sigma) \neq 0$. The initial bias is thus specified by a single parameter characterizing the expected strength of correlations. This parameter is related to the expected size of the total angular momentum

$$\vec{J} := \sum_{i=1}^M \vec{S}_i \qquad (37)$$

via

$$\langle \vec{J}^2 \rangle = 3M(M-1) \cdot c + 3M/4. \qquad (38)$$

Since the latter expectation value must lie in the range between zero and $M/2 \cdot (M/2 + 1)$, the strength of correlations is bounded by

$$-1/4(M-1) \leq c \leq 1/12. \qquad (39)$$

For four qubits, these bounds are $\pm 1/12$.

Given that the initial bias can be characterized completely by the expected magnitude of the total angular momentum, it is possible to write σ in Maxent form

$$\sigma = Z(\lambda(\sigma))^{-1} \exp[-\lambda(\sigma) \cdot \vec{J}^2] \qquad (40)$$

with properly adjusted Lagrange parameter $\lambda(\sigma)$ and partition function

$$Z(\lambda) := \operatorname{tr} \exp[-\lambda \cdot \vec{J}^2]. \qquad (41)$$

For four qubits this partition function can be calculated and all resulting thermodynamic equations solved analytically. Adding four spins 1/2 the total angular momentum squared can take the values $j(j+1)$ with $j = 0, 1, 2$. In the Clebsch-Gordan series these spin-j representations occur with respective multiplicities $m_0 = 2$, $m_1 = 3$ and $m_2 = 1$; and each such representation has dimension $2j + 1$ [42]. Therefore,

$$Z(\lambda) = \sum_{j=0}^2 m_j (2j+1) \exp[-\lambda j(j+1)]$$
$$= 2 + 9 \exp(-2\lambda) + 5 \exp(-6\lambda); \qquad (42)$$

whence follows the expectation value

$$\langle \vec{J}^2 \rangle = -\partial \ln Z / \partial \lambda = 6 \cdot \frac{3 \exp(-2\lambda) + 5 \exp(-6\lambda)}{2 + 9 \exp(-2\lambda) + 5 \exp(-6\lambda)}. \qquad (43)$$

Given the strength of correlations c one can now arrive at the associated λ in the following three steps: relate c to $\langle \vec{J}^2 \rangle$; then solve the thermodynamic equation (43) for $x := \exp(-2\lambda)$, a cubic equation with a unique analytical solution; and finally obtain $\lambda = -\ln x/2$. In my example I shall assume a prior bias $c(\sigma) = -0.02$ with associated $\langle \vec{J}^2 \rangle_\sigma = 2.28$, $x(\sigma) \approx 0.7$ and $\lambda(\sigma) \approx 0.18$, implying $Z(\lambda(\sigma)) \approx 10$.

Enter experimental data, measured on a sample of size $N = 10,000$. In order to keep calculations tractable analytically I make the (admittedly artificial) assumption that the $r = 48$ different sample means confirm the anticipated symmetries perfectly, and that hence the experimental data, too, can be characterized by the single parameter c. But the measured correlation strength differs from prior expectation, $c \neq c(\sigma)$; say, by 25%, $c = -0.025$. Associated with this experimental value is a Maxent state μ_c^σ of the same canonical form as σ, yet with different expectation and parameter values $\langle \vec{J}^2 \rangle = 2.1$, $x \approx 0.625$ and $\lambda \approx 0.235$, which implies $Z(\lambda) \approx 8.85$. This empirical Maxent state deviates from the initial bias by the distance

$$S(\mu_c^\sigma \| \sigma) = (\lambda(\sigma) - \lambda) \cdot \langle \vec{J}^2 \rangle + \ln Z(\lambda(\sigma)) - \ln Z(\lambda) \approx 0.005. \qquad (44)$$

One can now deduce $N_{\min} \approx 5,000$ and hence confirm that the two criteria for the applicability of the evidence procedure ($N > N_{\min}$ and $r \gg 1$) are indeed satisfied. From $N \approx 2N_{\min}$ results an optimal hyperparameter α_0 of approximately the same size as the sample, $\alpha_0 \approx N$. Hence the estimate (16) gives approximately equal weight to likelihood and prior; it features an interpolated Lagrange parameter $\lambda_e \approx 0.208$ which corresponds to $x_e \approx 0.66$, $\langle \vec{J}^2 \rangle_e \approx 2.19$ and $c_e \approx -0.0226$.

This example illustrates nicely the two distinguishing traits of the evidence procedure. First, the estimate is not based on experimental data alone; rather, it takes into account prior information encoded in the initial bias σ. In fact, in my example the initial bias and experimental data carried approximately equal weight, so the estimate c_e for the strength of correlations was approximately half way between its measured value and the prior expectation. Second, even though in my example the set of sample means was far from informationally complete, the resulting estimate was *just as good* as if it had been based on a set that was informationally complete. Indeed, since the experimental data were sufficiently close to prior expectations (in particular with regards to the anticipated symmetries) the hyperparameter took a value of similar magnitude as the sample size; and hence estimates pertaining to degrees of freedom that had not been measured became just as accurate as estimates pertaining to those that had been measured.

IV. CONCLUSION

In this paper I transferred the evidence procedure from its historical use in image reconstruction and other classical estimation tasks to a novel use in quantum state tomography. I verified that this transfer is justified whenever samples are large enough (greater than $N_{\min}$) and the set of measured sample means is sufficiently big ($r \gg 1$), albeit not necessarily informationally complete. The power of the evidence procedure lies in its optimal use of prior information: Whenever such prior information is available, the procedure promises to improve both the accuracy and the efficiency of quantum state estimation. Indeed, I demonstrated in the four-qubit example that the evidence procedure yields a state estimate that is modified with respect to the conventional maximum likelihood or Maxent estimate, by giving some credence

still to one's initial bias. The relative weight attributed to this initial bias is determined by the degree to which the experimental data meet prior expectations overall; in my example this degree was high because all anticipated symmetries were confirmed perfectly.

In addition to providing a state estimate the evidence procedure also quantifies the associated error bars; and it does so not only with respect to measured but also with respect to *un*measured degreees of freedom. Whenever the data are in good agreement with prior expectations overall – as was the case in my example – and hence the optimal hyperparameter α_0 is of similar magnitude as the sample size N, the unmeasured degrees of freedom are in fact estimated to an accuracy which is just as good as if they had been measured, too. In situations where the primary goal of an experiment is to confirm or incrementally amend well-founded prior expectations, therefore, quantum state estimation becomes very efficient: Measurement of a limited, not informationally complete set of observables will suffice to estimate the state at a level of confidence extending uniformly to all degrees of freedom, even the unmeasured ones. I have not yet considered experimental errors introduced by inaccuracies of the measurement apparatus. Their treatment should be straightforward conceptually, and will be the subject of future work.

Finally, I speculate whether the evidence procedure may also have some import on foundational issues in statistical mechanics. The thermodynamical description of macroscopic systems by means of (generalized) canonical distributions always represents a theoretical model which can never be verified in its entirety — after all, who would ever verify experimentally, say, the predicted 42-body correlations in a liquid? Nevertheless, we use thermodynamical models with a high degree of confidence. The evidence procedure might provide a justification for this: As long as the predictions of the thermodynamical model are verified for some limited set of observables, one has good reason to trust its predictions for all the other observables, too.

[1] G. M. D'Ariano, M. G. A. Paris, and M. F. Sacchi, Lect. Notes Phys., **649**, 7 (2004), doi:10.1007/b98673.

[2] G. M. D'Ariano, C. Macchiavello, and M. G. A. Paris, Phys. Rev. A, **50**, 4298 (1994), doi:10.1103/PhysRevA.50.4298.

[3] U. Leonhardt, H. Paul, and G. M. D'Ariano, Phys. Rev. A, **52**, 4899 (1995), doi:10.1103/PhysRevA.52.4899.

[4] Z. Hradil, Phys. Rev. A, **55**, R1561 (1997), doi:

10.1103/PhysRevA.55.R1561.

[5] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Phys. Rev. A, **64**, 052312 (2001), doi:10.1103/PhysRevA.64.052312.

[6] Z. Hradil, J. Rehacek, J. Fiurasek, and M. Jezek, Lect. Notes Phys., **649**, 59 (2004), doi:10.1007/b98673.

[7] V. Buzek, G. Drobny, G. Adam, R. Derka, and P. L. Knight, J. Mod. Opt., **44**, 2607 (1997), doi:

10.1080/09500349708231905.

[8] V. Buzek, R. Derka, G. Adam, and P. L. Knight, Ann. Phys., **266**, 454 (1998), doi:10.1006/aphy.1998.5802.

[9] V. Buzek, G. Drobny, R. Derka, G. Adam, and H. Wiedemann, Chaos Sol. Frac., **10**, 981 (1999), doi:10.1016/S0960-0779(98)00144-1.

[10] V. Buzek, Lect. Notes Phys., **649**, 189 (2004), doi:10.1007/b98673.

[11] R. Blume-Kohout, New J. Phys., **12**, 043034 (2010), doi:10.1088/1367-2630/12/4/043034.

[12] R. Blume-Kohout, "Hedged maximum likelihood estimation," (2010), arXiv:1001.2029v1 [quant-ph].

[13] J. Rehacek, D. Mogilevtsev, and Z. Hradil, New J. Phys., **10**, 043022 (2008), doi:10.1088/1367-2630/10/4/043022.

[14] K. M. R. Audenaert and S. Scheel, New J. Phys., **11**, 023028 (2009), doi:10.1088/1367-2630/11/2/023028.

[15] H. Häffner, W. Hänsel, C. F. Roos, J. Benhelm, D. Chek-al-kar, M. Chwalla, T. Körber, U. D. Rapol, M. Riebe, P. O. Schmidt, C. Becher, O. Gühne, W. Dür, and R. Blatt, Nature, **438**, 643 (2005), doi:10.1038/nature04279.

[16] S. Aaronson, Proc. R. Soc. A, **463**, 3089 (2007), doi:10.1098/rspa.2007.0113.

[17] K. M. R. Audenaert and M. B. Plenio, New J. Phys., **8**, 266 (2006), doi:10.1088/1367-2630/8/11/266.

[18] H. Wunderlich and M. B. Plenio, J. Mod. Opt., **56**, 2100 (2009), doi:10.1080/09500340903184303.

[19] J. Rehacek, Z. Hradil, Z. Bouchal, R. Celechovsky, I. Rigas, and L. L. Sanchez-Soto, Phys. Rev. Lett., **103**, 250402 (2009), doi:10.1103/PhysRevLett.103.250402.

[20] M. Cramer and M. B. Plenio, "Reconstructing quantum states efficiently," (2010), arXiv:1002.3780v2 [quant-ph].

[21] S. T. Flammia, D. Gross, S. D. Bartlett, and R. Somma, "Heralded polynomial-time quantum state tomography," (2010), arXiv:1002.3839v2 [quant-ph].

[22] O. Landon-Cardinal, Y.-K. Liu, and D. Poulin, "Efficient direct tomography for matrix product states," (2010), arXiv:1002.4632v1 [quant-ph].

[23] J. M. Pryce and A. D. Bruce, J. Phys. A: Math. Gen., **28**, 511 (1995), doi:10.1088/0305-4470/28/3/009.

[24] D. S. Sivia, *Data analysis: a Bayesian tutorial* (Oxford University Press, 1996).

[25] S. F. Gull, in *Maximum Entropy and Bayesian Methods*, edited by J. Skilling (Kluwer Academic, 1989) pp. 53–71.

[26] J. Skilling, D. R. T. Robinson, and S. F. Gull, in *Maximum Entropy and Bayesian Methods*, edited by W. T. Grandy and L. H. Shick (Kluwer Academic, 1991) pp. 365–368.

[27] J. Skilling and S. F. Gull, in *Spatial statistics and imaging*, Lecture Notes–Monograph Series, Vol. 20, edited by A. Possolo (Institute of Mathematical Statistics, 1991) pp. 341–367, doi:10.1214/lnms/1215460511.

[28] D. J. C. MacKay, Neural Comp., **4**, 415 (1992), doi:10.1162/neco.1992.4.3.415.

[29] C. E. M. Strauss, D. H. Wolpert, and D. R. Wolf, in *Maximum Entropy and Bayesian Methods*, edited by A. Mohammad-Djafari and G. Demoment (Kluwer Academic, 1993) pp. 113–120.

[30] K. M. Hanson, in *Medical Imaging: Image Processing*, Proc. SPIE, Vol. 1898, edited by M. H. Loew (1993) pp. 716–731, doi:10.1117/12.154577.

[31] D. J. C. MacKay, in *Maximum Entropy and Bayesian Methods*, edited by G. R. Heidbreder (Kluwer Academic, 1996) pp. 43–59.

[32] D. H. Wolpert and C. E. M. Strauss, in *Maximum Entropy and Bayesian Methods*, edited by G. R. Heidbreder (Kluwer Academic, 1996) pp. 61–78.

[33] D. J. C. MacKay, Neural Comp., **11**, 1035 (1999), doi:10.1162/089976699300016331.

[34] J. Skilling, in *Maximum Entropy and Bayesian Methods*, edited by G. J. Erickson and C. R. Smith (Kluwer Academic, 1988) pp. 173–187.

[35] F. Hiai and D. Petz, Comm. Math. Phys., **143**, 99 (1991), doi:10.1007/BF02100287.

[36] T. Ogawa and H. Nagaoka, IEEE Trans. Info. Theory, **46**, 2428 (2000), doi:10.1109/18.887855.

[37] D. Petz, *Quantum information theory and quantum statistics* (Springer, 2008).

[38] F. G. S. L. Brandão and M. B. Plenio, Comm. Math. Phys., **295**, 791 (2010), doi:10.1007/s00220-010-1005-z.

[39] M. B. Ruskai, Rep. Math. Phys., **26**, 143 (1988), doi:10.1016/0034-4877(88)90009-2.

[40] R. Balian, Y. Alhassid, and H. Reinhardt, Phys. Rep., **131**, 1 (1986), doi:10.1016/0370-1573(86)90005-0.

[41] J. Skilling, in *Maximum Entropy and Bayesian Methods*, edited by J. Skilling (Kluwer Academic, 1989) pp. 45–52.

[42] J. F. Cornwell, *Group theory in physics*, Vol. 2 (Academic Press, 1984).